doi:10.1006/jmbi.2001.5002 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (2001) 312, 1073–1088

JMB

Dimeric DNA Quadruplex Containing Major Groove-aligned A·T·A·T and G·C·G·C Tetrads Stabilized by Inter-subunit Watson-Crick A·T and G·C Pairs

Na Zhang†, Andrey Gorin†, Ananya Majumdar, Abdelali Kettani Natalya Chernichenko, Eugene Skripkin and Dinshaw J. Patel*

Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center New York, NY 10021, USA

*Corresponding author

We report on an NMR study of unlabeled and uniformly ^{13}C,^{15}N-labeled d(GAGCAGGT) sequence in 1 M NaCl solution, conditions under which it forms a head-to-head dimeric quadruplex containing sequentially stacked G·C·G·C, G·G·G·G and A·T·A·T tetrads. We have identified, for the first time, a slipped A·T·A·T tetrad alignment, involving recognition of Watson-Crick A·T pairs along the major groove edges of opposing adenine residues. Strikingly, both Watson-Crick G·C and A·T pairings within the direct G·C·G·C and slipped A·T·A·T tetrads, respectively, occur between rather than within hairpin subunits of the dimeric d(GAGCAGGT) quadruplex. The hairpin turns in the head-to-head dimeric quadruplex involve single adenine residues and adds to our knowledge of chain reversal involving edgewise loops in DNA quadruplexes. Our structural studies, together with those from other laboratories, definitively establish that DNA quadruplex formation is not restricted to G_n repeat sequences, with their characteristic stacked uniform G·G·G·G tetrad architectures. Rather, the quadruplex fold is a more versatile and robust architecture, accessible to a range of mixed sequences, with the potential to facilitate G·C·G·C and A·T·A·T tetrad through major and minor groove alignment, in addition to G·G·G·G tetrad formation. The definitive experimental identification of such major groove-aligned mixed A·T·A·T and G·C·G·C tetrads within a quadruplex scaffold, has important implications for the potential alignment of duplex segments during homologous recombination.

Keywords: A·T·A·T and G·C·G·C tetrads; dimeric DNA quadruplex; hydrogen bond alignments; inter-subunit Watson-Crick pairs

Introduction

There is an increasing appreciation of the role DNA quadruplexes (reviews[1–4]) may play in biological processes ranging from replication, transcription and recombination (review[5]) to telomere function (review[4]). The earliest research focused on quadruplex formation involving stacked G·G·G·G tetrads (review[6]), with polymorphism introduced by the directionality of adjacent strands around the quadruplex (review[4]).

Formation of mixed tetrads could dramatically increase the versatility of quadruplex formation by allowing adoption of this four-stranded architecture by sequences other than simple G_n repeats. Towards this end, efforts have been made to identify and characterize G·C·G·C tetrads, where a pair of Watson-Crick G·C pairs can potentially align either through their major groove or their minor groove edges. This approach was successful in the case of the Fragile X syndrome triplet repeat-containing d(GCGGT$_3$GCGG) sequence, which dimerizes in solution through head-to-tail alignment of hairpins. The resulting quadruplex contains G·C·G·C tetrads, through major groove alignment of Watson-Crick G·C pairs (Figure 1(a)).[7]

†These authors contributed equally to the research.

Abbreviations used: NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; ppm, parts per million; HSQC, heteronuclear single quantum coherence; COSY, correlated spectroscopy; TOCSY, total COSY.

E-mail address of the corresponding author: pateld@mskcc.org

0022-2836/01/051073–16 $35.00/0

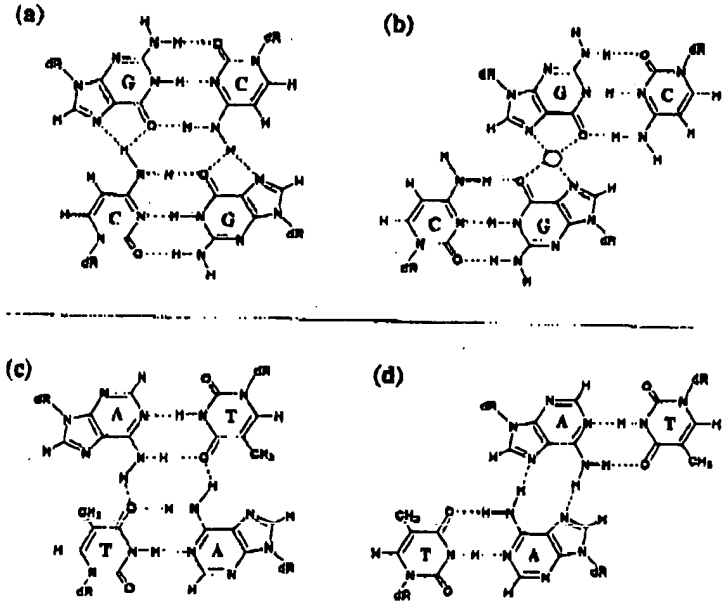

Figure 1. Potential G·C·G·C tetrad formation involving either (a) direct or (b) slipped alignment of major groove edges of a pair of Watson-Crick G·C pairs. The slipped alignment requires a monovalent cation to coordinate the inwardly oriented acceptor atoms. Potential A·T·A·T tetrad formation involving either (a) direct or (b) slipped alignment of major groove edges of a pair of Watson-Crick A·T pairs.

Independent structural studies have identified minor groove-aligned G·C·G·C tetrads (Supplementary Material, Figure S1(a)) in the crystalline[8] and solution states.[9]

More recent studies have demonstrated that G·C·G·C tetrads aligned through their major groove edges can switch between two distinct alignment geometries. In the direct alignment (Figure 1(a)), the two G·C pairs align directly opposite each other, resulting in the acceptor atoms (O6 and N7) of G pairing with the donor group (NH_2-4) of C. Alternately, in the slipped alignment (Figure 1(b)), the major groove edges of two guanine bases are positioned opposite each other. In this case, a monovalent cation (not necessarily in the plane of the tetrad) is needed to coordinate the four inwardly pointing nitrogen and oxygen acceptor atoms on the guanine bases (Figure 1(b)). Solution structural studies of the adeno associated viral repeat-containing d(GGGCT$_4$GGGC) sequence have demonstrated that the direct G·C·G·C tetrad alignment (Figure 1(a)) is formed in Na cation solution,[10] while the slipped G·C·G·C tetrad alignment (Figure 1(b)) is formed in K cation solution.[11] Such conformational transitions were unanticipated, and attest to the diversity of pairing geometries in mixed tetrad alignments and the role of cations in modulating this transition.

The major groove-aligned G·C·G·C tetrad has now been observed in a range of DNA quadruplexes[7,10,12] and appears to be a robust tetrad motif adopted by a range of DNA sequences. One can therefore anticipate potential formati n of the corresponding major-groove aligned A·T·A·T counterpart, either of the direct (Figure 1(c)) or slipped (Figure 1(d)) alignment type. By contrast, the corresponding minor groove-aligned A·T·A·T tetrad, of the slipped cation-coordination type (Supplementary Material, Figure S1(b)), has been reported earlier both in the crystalline[13] and solution[9] states.

Demonstration of the major groove-aligned A·T·A·T tetrad formation has turned out to be a considerable challenge, and after many unsuccessful attempts, our group has now identified a sequence, d(GAGCAGGT), which forms a head-to-head dimeric quadruplex (Figure 2a and Supplementary Material, Figure S2), stabilized by direct G·C·G·C (Figure 2(b)) and slipped A·T·A·T (Figure 2(d)) tetrads, flanking a central G·G·G·G tetrad (Figure 2(c)). Strikingly, we observe Watson-Crick G·C pairing between monomer subunits within the G·C·G·C tetrad and Watson-Crick A·T pairing between monomer subunits within the A·T·A·T tetrad in this quadruplex. Finally, the edgewise turns in the head-to-head dimeric quadruplex involve a single adenine residue and provide new insights into chain reversal in DNA quadruplexes.

Results

Our group scanned a range of sequences containing G, A and T residues in our attempts to generate a major groove-aligned A·T·A·T tetrad, stabilized through stacking on a G·G·G·G tetrad within a quadruplex scaffold. This approach has worked in the past in our laboratory when we have generated stable triads[15–18] and A·(G·G·G·G)·A hexads,[12,19] stabilized through stacking on G·G·G·G tetrads within a quadruplex

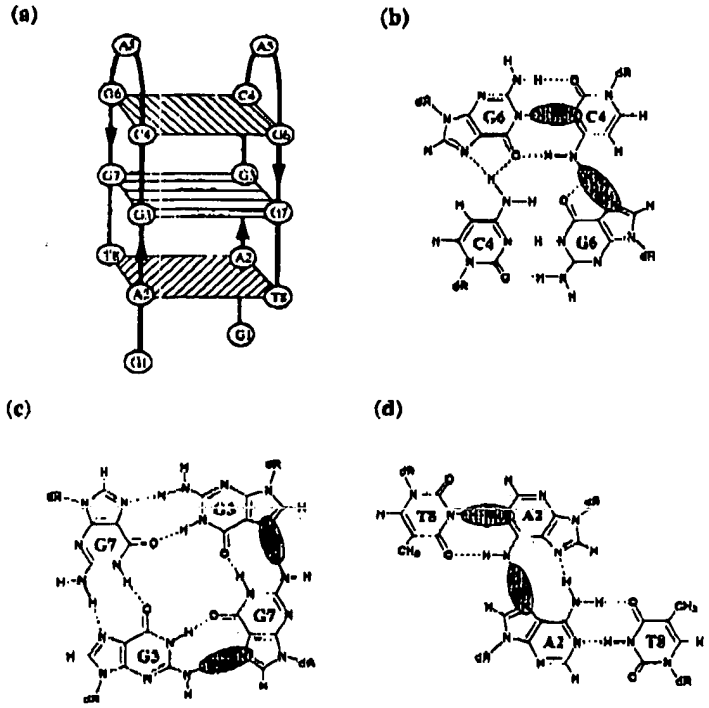

Figure 2. (a) Folding topology of a head-to-head dimeric d(GAGCAGGT) quadruplex in 1 M NaCl. The backbone tracing of the individual strands are shown by thick lines and the chain directionality indicated by arrows. Pairing alignments for (b) the G6·C4·G6·C4 tetrad, (c) the G3·G7·G3·G7 tetrad and (d) the A2·T8·A2·T8 tetrad.

scaffold. We report on NMR studies of the d(GAGCAGGT) sequence, which forms a dimeric quadruplex containing a major groove-aligned slipped A·T·A·T (Figure 2(c)) tetrad.

Imino proton spectra and assignments

The exchangeable proton spectrum (5.5 to 14.5 ppm) of the d(GAGCAGGT) sequence was highly dependent on added NaCl concentration. At low NaCl concentration, the predominant conformation gave very broad resonances in slow exchange with a minor component that gave narrow resonances (Supplementary Material, Figure S3). The equilibrium shifted with increasing NaCl to the conformation that exhibited narrow resonances, resulting in the spectrum shown in Figure 3(a) in 1 M NaCl, 2 mM phosphate buffer (pH 6.6) at 10 °C. This spectrum exhibits well resolved exchangeable (8.0 to 15.0 ppm) and non-exchangeable (7.0 to 8.7 ppm) resonances, with the total number of imino, amino and base proton resonances consistent with formation of a single conformer in solution. We observe two narrow imino protons between 11 and 12 ppm, a region characteristic of N-H···O hydrogen bonds,[4] and narrow resonances at 13.11 ppm and 14.28 ppm, a region characteristic of N-H···N hydrogen bonds[4] (Figure 3(a)). We also observe two narrow amino proton resonances between 8.0 and 8.5 ppm and two additional narrow amino proton resonances downfield-shifted between 9.0 and 9.2 ppm (Figure 3(a)).

The exchangeable imino and amino protons have been assigned following analysis of two-dimensional data sets on unlabeled and uniformly ^{13}C,^{15}N-labeled d(GAGCAGGT) sequence. An expanded 60 ms mixing time NOESY contour plot of d(GAGCAGGT) in 1 M NaCl at 0 °C is plotted in Figure 3(b). The NOE cross-peaks are labeled in the Figure and the assignments are given in the legend. The corresponding NOESY data set recorded at a longer mixing time of 200 ms is plotted in Supplementary Material, Figure S4, and the cross-peaks assignments listed in the legend. We could readily distinguish thymine from guanine imino protons because of their distinct nitrogen chemical shifts in a ^{1}H-^{15}N HSQC spectrum (Supplementary Material, Figure S5).

We establish formation of an A2·T8 Watson-Crick base-pair based on NOEs between the imino proton of T8 and the amino (peaks a and a', Figure 3(b)) and H2 (peak b, Figure 3(b)) protons of A2. In addition, we unexpectedly observe NOEs between the H8 and NH_2 protons of A2 (peaks n and n', Figure 3(b)), consistent with formation of a major groove-aligned slipped A2·T8·A2·T8 tetrad, schematically outlined in Figure 2(d).

We establish formation of a G6·C4 Watson-Crick base-pair based on NOEs between the imino proton of G6 and the amino protons of C4 (peaks c

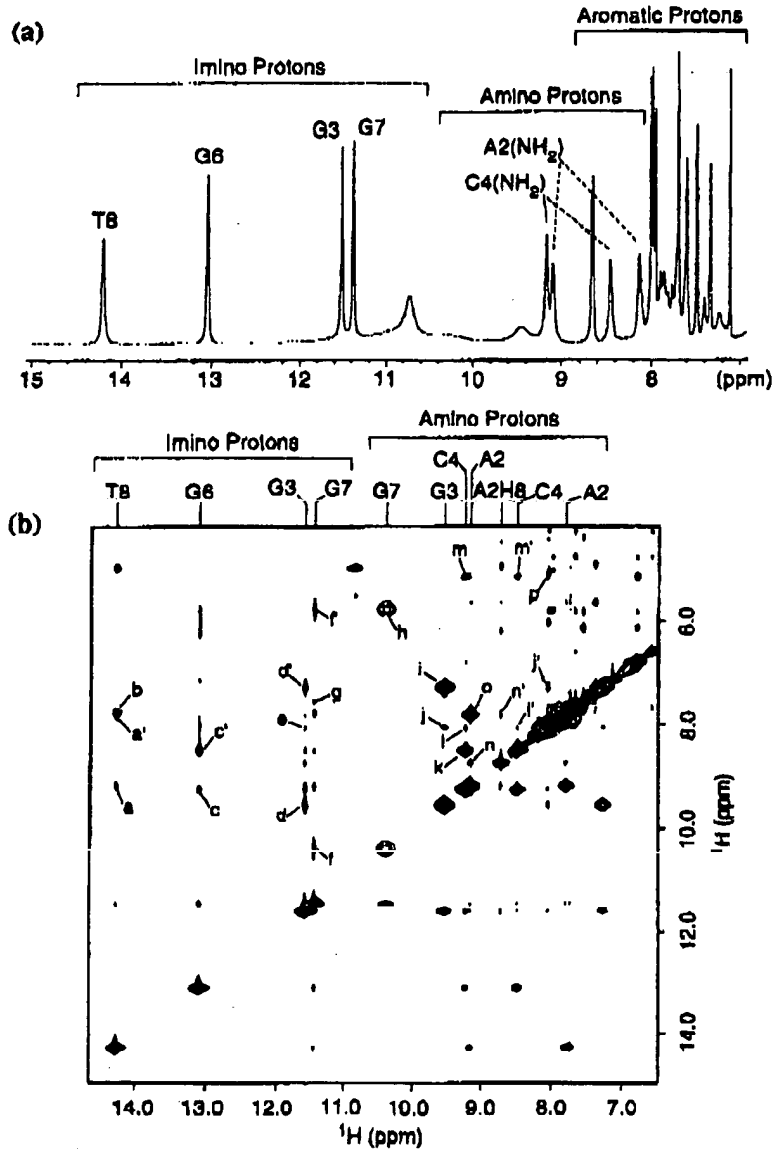

Figure 3. (a) Proton NMR spectrum (7.0 to 15.0 ppm) of the d(GAGCAGGT) quadruplex (8.8 mM in strands) in 1 M NaCl, 2 mM phosphate in H_2O (pH 6.6) at 10°C. (b) Expanded NOESY (60 ms mixing time) contour plot correlating NOEs between imino, amino and non-exchangeable protons in the d(GAGCAGGT) quadruplex in 1 M NaCl, 2 mM phosphate, H_2O (pH 6.6) at 0°C. The NOE cross-peaks a to p are assigned as follows: a and a', T8(NH3)-A2(NH_2-6); b, T8(NH3)-A2(H2); c and c', G6(NH1)-C4(NH_2-4); d and d', G3(NH1)-G3(NH_2-2); e, G3(NH1)-G7(H8); f and f', G7(NH1)-G7(NH_2-2); g (observable in 200 ms mixing time NOESY experiment), G7(NH1)-G3(H8); h, G7(NH_2-2)-G7(NH_2-2); i, G3(NH_2-2)-G3(NH_2-2); j and j', G3(NH_2-2)-G7(H8); k, C4(NH_2-4)-C4(NH_2-4); l and l', C4(NH_2-4)-G6(H8); m and m', C4(NH_2-4)-C4(H5); n and n', A2(NH_2-6)-A2(H8); o, A2(NH_2-6)-A2(NH_2-6); p, G6(H8)-C4(H5).

and c', Figure 3(b)). We also detect NOEs between the amino protons of C4 and the H8 proton of G6 (peaks l and l', Figure 3(b)), and between the H5 proton of C4 and the H8 proton of G6 (peak p, Figure 3(b)), consistent with formation of a major groove-aligned direct G6·C4·G6·C4 tetrad, schematically outlined in Figure 2(b).

The narrow G3 and G7 imino protons exhibit a set of NOEs indicative of G3·G7·G3·G7 tetrad formation, schematically outlined in Figure 2(c).

These include NOEs between the H8 proton of G7 with the imino (peak e, Figure 3(b)) and amino (peak j and j', Figure 3(b)) protons of G3, and an NOE between the H8 proton of G3 and the imino proton of G7 (box g, Figure 3(b); observable in 200 ms NOESY spectrum, peak g, Supplementary Material, Figure S4).

The NOE data outlined above for the d(GAGCAGGT) sequence are consistent with formation of slipped A2·T8·A2·T8 (Figure 2(d)), direct G6·C4·G6·C4 (Figure 2(b)) and G3·G7·G3·G7 (Figure 2c) tetrads. This can be most readily achieved by head-to-head dimerization to form a quadruplex as shown schematically in Figure 2(a) and Supplementary Material, Figure S2. The exchangeable imino and amino proton chemical shifts of the d(GAGCAGGT) quadruplex are listed in Supplementary Material, Table S1.

Non-exchangeable proton spectra and assignments

The non-exchangeable base and sugar protons have been assigned following analysis of two-dimensional data sets on unlabeled and uniformly $^{13}C,^{15}N$-labeled d(GAGCAGGT) sequence. An expanded 250 ms mixing time NOESY contour plot of d(GAGCAGGT) quadruplex in 1 M NaCl, ^2H_2O buffer at 10 °C is plotted in Figure 4(a), along with a tracing of distance connectivities between base protons and their own and 5'-flanking sugar H1' protons from G1 to T8 in the sequence. The G3 residue adopts a *syn* alignment based on the strong H8 to its own H1' NOE in a 50 ms mixing NOESY stacked plot (Figure 4(b)). The remaining sugar proton chemical shifts were obtained from an analysis of other regions of the through space NOESY contour plot and through bond COSY and TOCSY correlations of the assigned sugar H1' protons with the remaining sugar protons within individual rings. The non-exchangeable base and sugar proton chemical shifts of the d(GAGCAGGT) quadruplex are listed in Supplementary Material, Table S2. The H8 proton of A2 (8.71 ppm) is downfield-shifted while the H6 (6.78 ppm) and H5 (5.15 ppm) protons of C4 are somewhat upfield-shifted.

We can also correlate exchangeable imino protons with their own H8 protons within individual guanine bases, *via* through bond correlations to their C5 ring carbon atoms.[20] Such a correlation experiment on the sample of uniformly $^{13}C,^{15}N$-labeled d(GAGCAGGT) sequence is shown in Figure 4(c).

Hydrogen bond alignments

We have verified the formation of NOE-based G6·C4·G6·C4 (Figure 2(b)), G3·G7·G3·G7 (Figure 2(c)) and A2·T8·A2·T8 (Figure 2(d)) tetrad alignments, by identifying through bond coupling connectivities across N-H··N hydrogen bonds[21–23] within the folded architecture of the uniformly $^{13}C,^{15}N$-labeled d(GAGCAGGT) quadruplex in 1 M NaCl, H_2O buffer at 0 °C. The amino protons of A2, G3 and C4 could be correlated with their directly attached nitrogen atoms using this labeled sample as shown in Supplementary Material, Figure S6.

We observe a N-H··N hydrogen bond connectivity between the N1H donor of G6 and the N3 acceptor of C4 (peak 4, Figure 5(a)) in a HNN-COSY experiment[24] recorded on the d(GAGCAGGT) quadruplex. This connectivity provides direct support for a Watson-Crick G6·C4 alignment (observed couplings as shaded region in Figure 2(b)). In addition, we observe HNN-COSY coupling connectivities between the NH_2 protons of C4 and the N7 of G6 (peaks 1 and 1', Figure 5(b)) and four bond H(CN)N(H) coupling connectivities[24] between the N4 of C4 and the H8 of G6 (peak 3, Figure 5(c)). These observations verify the direct alignment of opposing Watson-Crick G6·C4 pairs through their major groove edges (observed coupling as shaded region in Figure 2(b)), to form the G6·C4·G6·C4 tetrad outlined in Figure 2(b).

The observed H(CN)N(H) coupling connectivities[24] between the H8 proton of G7 and the N2 of G3 (peak 1, Figure 5(c)) and between the H8 proton of G3 and the N2 of G7 (peak 2, Figure 5(c)), verify the proposed hydrogen bonding patterns (shaded regions in Figure 2(c)) around the G3·G7·G3·G7 tetrad.

We observe a N-H··N hydrogen bond connectivity between the N3H donor of T8 and the N1 acceptor of A2 (peak 2, Figure 5(a)) in a HNN-COSY experiment. This connectivity provides direct support for a Watson-Crick A2·T8 alignment. In addition, we observe HNN-COSY coupling connectivities between the NH_2 protons of A2 and the N7 of A2 (peaks 2 and 2', Figure 5(b)). This connectivity could be either within an adenine or between adenine bases positioned opposite each other and hydrogen-bonded through their major groove edges, as shown for the A2·T8·A2·T8 tetrad alignment in Figure 2(d). We do not observe the corresponding coupling connectivities between the NH_2 and N7 of A5, and hence we favor the latter explanation over the former. These observations verify the slipped alignment of opposing Watson-Crick A2·T8 pairs through their major groove edges, to form the A2·T8·A2·T8 tetrad outlined in Figure 2(d).

Intra-strand *versus* inter-strand NOE restraints

The 2-fold symmetry in the dimeric d(GAGCAGGT) quadruplex fold makes it critical to unambiguously differentiate between intra-strand and inter-strand contributions for key NOEs that define the folding topology in solution. We prepared a sample containing an equimolar mixture of unlabeled and uniformly $^{13}C,^{15}N$-labeled d(GAGCAGGT) sequences and recorded ^{15}N-edited (ω_1), $^{13}C,^{15}N$-purged (ω_2) NOESY[26b] (100 ms mixing time) spectra in 1 M NaCl-containing H_2O

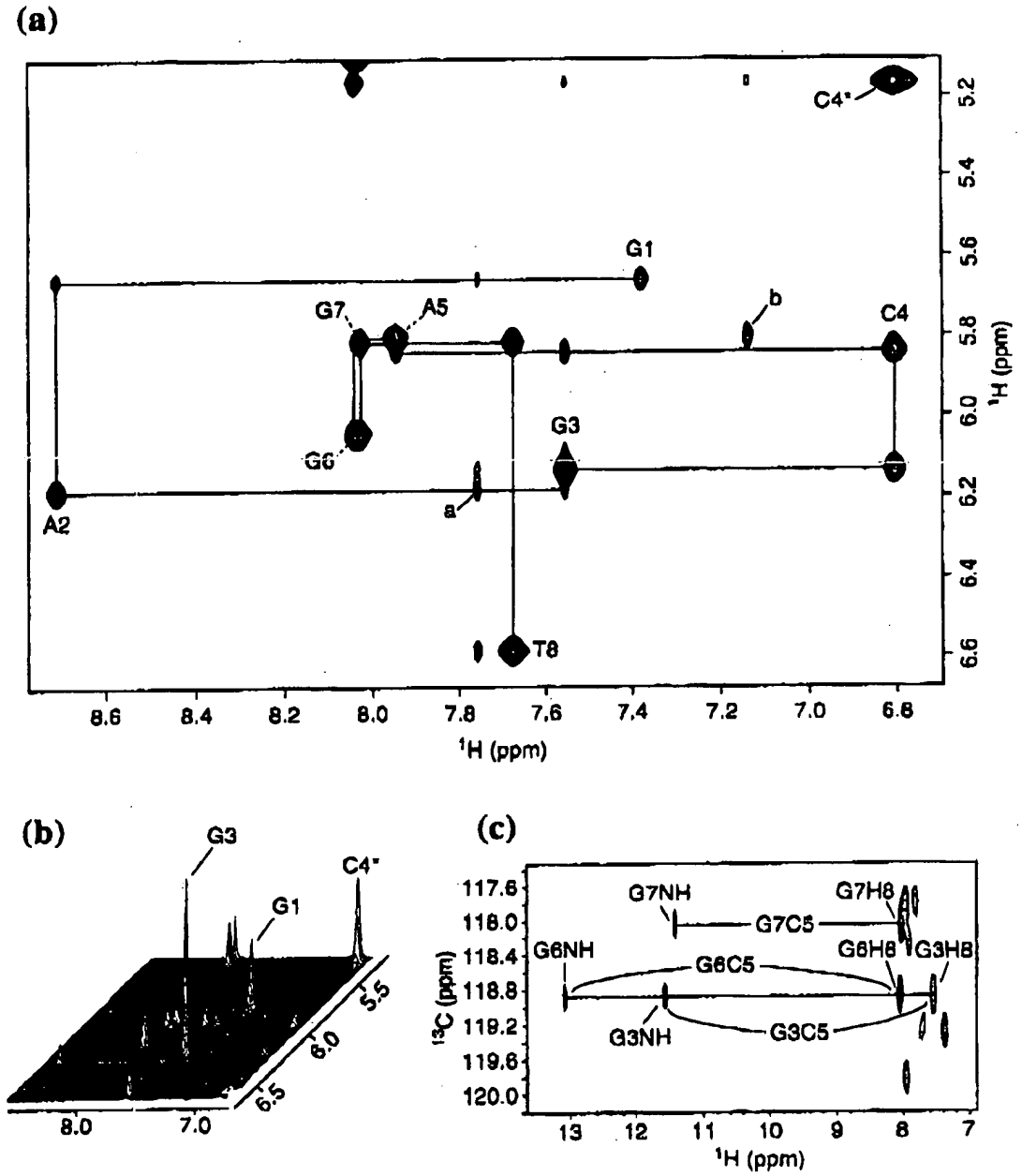

Figure 4. (a) Expanded NOESY (250 ms mixing time) contour plot correlating NOEs between base and sugar H-1′ protons in the d(GAGCAGGT) quadruplex (8.8 mM in strands) in 1 M NaCl, 2 mM phosphate, ^2H_2O (pH 6.6) at 10 °C. The cytidine H6 and H5 NOEs are designated by asterisks. The NOE cross-peaks a and b are assigned as follows: a, A2(H2)-A2(H1′); b, A5(H2)-A5(H1′). (b) Expanded NOESY (50 ms mixing time) stacked plot correlating NOEs between base and sugar H-1′ protons in the d(GAGCAGGT) quadruplex (8.8 mM in strands) in 1 M NaCl, 2 mM phosphate, ^2H_2O (pH 6.6) at 0 °C. The cytidine H6 and H5 NOEs are designated by asterisks. (c) Correlation of the guanine imino and H8 protons by through-bond connectivities to the C5 carbon within individual guanine rings in the uniformly ^{13}C,^{15}N-labled d(GAGCAGGT) quadruplex (5.3 mM in strands) in 1 M NaCl, 2 mM phosphate, ^2H_2O (pH 6.6) at 0 °C.

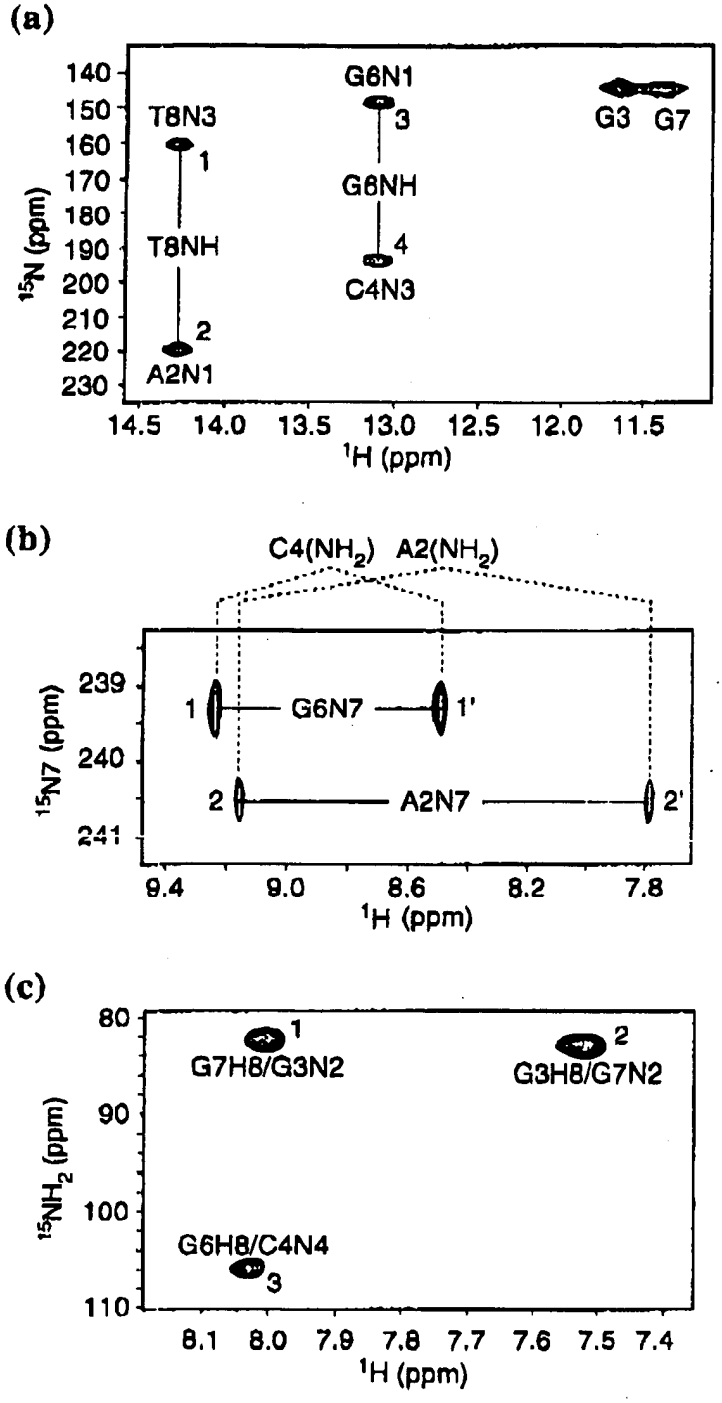

Figure 5. Identification of N-H···N hydrogen bond alignments in uniformly ^{13}C, ^{15}N-labeled d(GAGCAGGT) (5.3 mM in strands) in 1 M NaCl, 2 mM phosphate, H_2O (pH 6.6) at 0°C. Expanded $^{2}J_{NN}$ HNN-COSY contour plots correlating two-bond coupling connectivities between donor and acceptor nitrogen atoms within N-H···N hydrogen-bond alignments across base-pairs. (a) Two-bond coupling connectivities between T8(N3) imino donor and A2(N1) acceptor nitrogen atoms (peak 2) across the A2(N1)·T8(N3 donor) Watson-Crick base-pair and between G6(N1) imino donor and C4(N3) acceptor nitrogen atoms (peak 4) across the C4(N3)·G6(N1 donor) Watson-Crick base-pair. (b) Two-bond coupling connectivities between C4(N4) amino donor and G6(N7) acceptor nitrogen atoms (peaks 1,1') across C4(N4 amino)·G6(N7) mismatch pair and between A2(N6) amino donor and A2(N7) acceptor nitrogen atoms (peaks 2,2') across A2(N6 amino)·A2(N7) mismatch pair. (c) H(CN)N(H) spectrum showing inter-nucleotide H8(ω_2)-N2(ω_1) and H8(ω_2)-N4(ω_1) cross-peaks. The spectrum consisted of 608 (t_2) * 60 (t_1) complex points, with 160 transients per FID. Spectral widths of 2000 Hz ($t_1$max: 30 ms) and 8000 Hz ($t_2$max: 76 ms) were used, with a relaxation delay of two seconds, resulting in a total acquisition of 12 hours. Coupling connectivities are observed between the H8 of G7 and the N2 of G3 (peak 1), between the H8 of G3 and the N2 of G7 (peak 2), as well as between the H8 of G6 and N4 of C4 (peak 3). A coupling connectivity was not detectable between the H8 of A2 and the N6 of A2.

buffer solution at 0°C, to identify inter-strand NOEs and differentiate them from their intra-strand counterparts, for the 50% component in the mixture where the quadruplex contains one unlabeled and one uniformly labeled strand. We observe inter-strand NOEs between the imino pro-

ton of T8 and the amino (peaks 1 and 1', Figure 6(a); peaks 1 and 1', Figure 6(b)) and H2 (peak 2, Figure 6(b)) protons of A2, establishing formation of Watson-Crick A2·T8 pairs between hairpin subunits in the dimeric d(GAGCAGGT) quadruplex. We also observe inter-strand NOEs between the imino proton of G6 and the amino protons of C4 (peaks 2 and 2', Figure 6(a); peaks 3 and 3', Figure 6(b)), establishing formation of Watson-Crick G6·C4 pairs between hairpin subunits in the dimeric d(GAGCAGGT) quadruplex.

We also observe inter-strand NOEs between the amino protons of A2 and the H8 proton of A2 (peaks 4 and 4', Figure 6(a)), putting restraints on the relative alignments of the A·T base-pairs around the A·T·A·T tetrad. Interstrand NOEs are also observed between the imino proton of G7 and the H2 proton of A2 (peak 4, Figure 6(b)), putting restraints on the relative stacking of adjacent G3·G7·G3·G7 and A2·T8·A2·T8 tetrads in the quadruplex.

Distance restraints and molecular dynamics calculations

Distance restraints associated with exchangeable protons (total of 115) were qualitatively deduced from NOESY experiments in H_2O at two mixing times, while those associated with their non-exchangeable proton counterparts (total of 178) were quantified from NOE buildup curves in ^2H_2O at four mixing times, as outlined in Materials and Methods. The observation of a single set of narrow resonances for the d(GAGCAGGT) sequence at temperatures down to 0 °C, was consistent with formation of a dimeric quadruplex, containing two strands related by a 2-fold symmetry axis. Therefore, non-crystallographic symmetry restraints were used during the computations. All distance restraints were classified as ambiguous during the distance-restrained molecular dynamics computations. Experimentally defined hydrogen bonding alignments from NOE patterns on unlabeled sample, N-H···N scalar couplings on isotopically labeled sample, and intermolecular NOEs on an equimolar mixture of labeled and unlabeled sample, were used to restrain the G6·C4·G6·C4 (Figure 2(b)), G3·G7·G3·G7 (Figure 2(c)) and A2·T8·A2·T8 (Figure 2(d)) tetrads, with the folding models retaining these hydrogen bonding alignments during the computations.

The solution structure of the dimeric d(GAGCAGGT) quadruplex in 1 M NaCl was solved by molecular dynamics computations guided by hydrogen bonding and NOE distance restraints. Sixty starting structures were generated for the d(GAGCAGGT) 8-mer segment as sets of pairs of randomized chains separated by space intervals of 50 Å. The protocol outlined in Materials and Methods involved initial torsion space dynamics at 20,000 K followed by Cartesian space dynamics at 300 K. A subset of ten distance-refined structures of the d(GAGCAGGT) quadruplex were identified based on a combination of low NOE energies and fewest NOE violations.

Intensity restraints and NOE back calculations

The subset of ten converged distance-refined structures were next refined against the non-exchangeable proton NOE intensities associated with NOESY spectra recorded at four mixing times. These computations utilized a molecular dynamics with back calculation protocol outlined in Materials and Methods. The NOE violations, deviations from covalent geometry and pairwise r.m.s.d. values for the ten lowest energy intensity-refined structures of the d(GAGCAGGT) quadruplex (less the poorly defined G1 residues) are listed in Table 1.

Structural features

A stereo view of the ten superpositioned lowest energy intensity refined structures of the d(GAGCAGGT) quadruplex (less the poorly defined G1 residues) is shown in Figure 7(a). The sugar-phosphate backbone of individual symmetry-related hairpins are colored in orange and green, with the sequentially stacked loop A5 residue, G6·C4·G6·C4 tetrad, G3·G7·G3·G7 tetrad and A2·T8·A2·T8 tetrad colored in white, magenta, yellow and cyan, respectively. Ribbon and surface views of one representative refined structure of the d(GAGCAGGT) quadruplex using the GRASP program are plotted in Figure 7(b) and (c), respectively.

A stick representation of one symmetric strand of the d(GAGCAGGT) quadruplex (less the poorly defined G1 residues) is shown in Figure 8(a). The pairing alignments of the G6·C4·G6·C4 tetrad and A2·T8·A2·T8 tetrad in this representative structure are shown in Figure 8(b) and (c), respectively.

The stacking geometries between the A5 bases (in white) with the G6·C4·G6·C4 tetrad (in magenta) is shown in Figure 9(a), while that between the G6·C4·G6·C4 tetrad (in magenta) and the G3·G7·G3·G7 tetrad (in yellow) is shown in Figure 9(b). The overlap geometry between the G3·G7·G3·G7 tetrad (in yellow) and the A2·T8·A2·T8 tetrad (in cyan) is shown in Figure 9(c).

Discussion

The NMR-based structural characterization of the d(GAGCAGGT) quadruplex was considerably aided by the narrow and well-resolved resonances of the one and two-dimensional proton NMR spectra, and the availability of uniformly ^{13}C,^{15}N-labeled sequence, which aided greatly in the resonance assignment and identification of hydrogen bonding alignments. The computations were also aided by our ability to distinguish between inter-strand and intra-strand NOEs (and in turn, Watson-Crick hydrogen bonding alignments) using

Table 1. NMR restraints and structural statistics for the intensity refined structures of the dimeric d(GAGCAGGT) quadruplex

	Non-exchangeable	Exchangeable
A. *NMR restraints*		
Distance restraints[a]	86	19
Intra-residue distance restraints	76	45
Sequential (i, $i+1$) distance restraints	16	51
Long range $\geq(i, i+2)$ distance restraints		
Other restraints		
Hydrogen bonding restraints[b]	48	
Torsion angle restraints[c]	14	
Intensity restraints[a]		
Non-exchangeable protons (each of 4 mixing times)	178	
B. *Structural statistics in complex following intensity refinement*		
NOE violations		
Number >0.2 Å	6.1±0.7	
Maximum violations	0.27±0.03	
r.m.s.d. of violations	0.056±0.02	
NMR R factor ($R_{1/6}$)	0.074±0.003	
Deviations from ideal covalent geometry		
Bond lengths	0.011±0.001	
Bond angles	3.70±0.09	
Impropers	0.40±0.03	
Pairwise all heavy atom r.m.s.d. values (10 refined strucutres)		
All heavy atoms excluding G1	0.78±0.17	

[a] All distance restraints were set as ambiguous between intra and inter-residue contributions.
[b] These hydrogen bonding restraints are based on experimental NOE and $^3J_{NN}$ coupling data.
[c] Residues A2, C4, A5, G6, G7 and T8 were restrained to χ values in the 210(±40)° range, characteristic of *anti* glycosidic torsion angles, while residues G3 were restrained to χ values in the 65(±40)° range, characteristic of *syn* glycosidic torsion angles, identified experimentally.

equimolar mixtures of labeled and unlabeled samples of the quadruplex.

Quadruplex architecture

The experimental data on the d(GAGCAGGT) sequence is consistent with formation a dimeric quadruplex in 1 M NaCl solution. Individual strands of d(GAGCAGGT) form hairpins, with a single adenine, A5, involved in chain reversal. Dimerization involves head-to-head orientation of hairpins (Figures 2(a) and 7; Supplementary Material, Figure S1), with individual strands running antiparallel relative to their neighbors around the quadruplex (Figure 2(a)).

G·G·G·G tetrad

We detect a G3(*syn*)·G7(*anti*)·G3(*syn*)·G7(*anti*) alignment around the G·G·G·G tetrad in the dimeric d(GAGCAGGT) quadruplex. Such an alignment about the G·G·G·G tetrad is consistent with antiparallel alignment of adjacent strands around the quadruplex, and has been reported for the solution structures of the thrombin-binding DNA aptamer quadruplex,[27,28] as well as the d(GCGGT$_3$GCGG) Fragile X syndrome-containing sequence quadruplex[7] and the d(GGGCT$_4$GGGC) adeno-associated virus-containing sequence quadruplex.[10]

G·C·G·C tetrad

The G6·C4·G6·C4 tetrad (Figures 2(b) and 8(b)) is defined by all *anti* glycosidic torsion angles, consistent with antiparallel arrangement of adjacent strands around the dimeric d(GAGCAGGT) quadruplex. This type of direct G·C·G·C tetrad alignment (Figure 1(a)) has been observed previously for the solution structures of the d(GCGGT$_3$GCGG) Fragile X syndrome sequence-containing quadruplex[7] and the d(GGGCT$_4$GGGC) adeno-associated virus sequence-containing quadruplex.[10] There is, however, a critical distinction in that Watson-Crick G6·C4 pairs are formed between rather than within hairpin subunits (Figure 8(b)) in the head-to-head dimeric d(GAGCAGGT) quadruplex reported here. This contrasts with Watson-Crick G·C pairs within hairpin subunits in the head-to-tail dimeric d(GCGGT$_3$GCGG)[7] and d(GGGCT$_4$GGGC)[10] quadruplexes, reported previously.

The amino protons of C4 in the d(GAGCAGGT) quadruplex resonate at 8.49 ppm and 9.23 ppm. The chemical shift of 8.49 ppm is characteristic of a cytosine amino proton hydrogen-bonded to a carbonyl oxygen atom, such as would occur with a guanine residue in a Watson-Crick G·C pair. The chemical shift of 9.23 ppm is consistent with the other cytosine amino proton also being hydrogen bonded, and its much larger downfield shift may be reflective of its acceptors being both a ring nitrogen and carbonyl oxygen, such as would occur in a direct G·C·G·C tetrad (Figure 1(a)).

A·T·A·T tetrad

The slipped A2·T8·A2·T8 tetrad (Figures 2(d) and 8(c)) contains all *anti* glycosidic torsion angles, with hydrogen bonding along the major groove

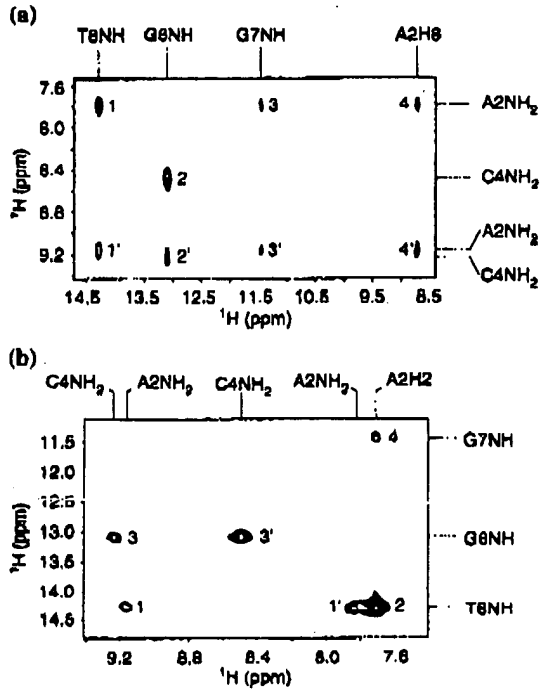

Figure 6. Identification of inter-strand NOEs. An expanded ^{15}N-edited (ω_1), ^{13}C, ^{15}N-purged (ω_2) NOESY (100 ms mixing time) contour plot of a 1:1 mixture of unlabeled and uniformly ^{13}C, ^{15}N-labeled d(GAGCAGGT) (3.3 mM in strands) in 1 M NaCl, 2 mM phosphate, H_2O (pH 6.6) at 0 °C. (a) Carrier is on amino-^{15}N. The relevant cross-peaks identifying inter-molecular NOEs are assigned as follows: 1,1′, T8(NH3)-A2(NH_2-6); 2,2′, G6(NH1)-C4(NH_2-4); 3,3′, G7(NH1)-A2(NH_2-6); and 4,4′, A2(H8)-A2(NH_2-6). (b) Carrier is on imino-^{15}N. The relevant cross-peaks identifying inter-molecular NOEs are assigned as follows: 1,1′, T8(NH3)-A2(NH_2-6); 2, T8(NH3)-A2(H2); 3,3′, G6(NH1)-C4(NH_2-4); 4, G7(NH1)-A2(H2).

edges of opposing adenine bases. This is the first report of any type (direct or slipped) of an A·T·A·T tetrad alignment, which involves pairing along the major groove edges. This A·T·A·T pairing alignment brings the H8 proton of one adenine in close proximity to the NH_2 proton of its adenine partner (Figure 1(d)), consistent with the observed intermolecular NOE between these pair of protons in mixed labeling experiments (peak 4 and 4′, Figure 6(a)). Equally important, is our demonstration that Watson-Crick A2·T8 pairs are formed between rather than within hairpin subunits (Figure 8(c)) in the head-to-head dimeric d(GAGCAGGT) quadruplex.

The amino protons of A2 in the d(GAGCAGGT) quadruplex resonate at 7.80 ppm and 9.16 ppm. The chemical shift of 7.80 ppm is characteristic of an adenine amino proton hydrogen-bonded to a carbonyl oxygen atom, such as would occur with a thymine residue in a Watson-Crick A·T pair. The chemical shift of 9.16 ppm is consistent with the other adenine amino proton also being hydrogen bonded, and its much larger downfield shift is reflective of its acceptor being a ring nitrogen, such as would occur in a slipped A·T·A·T tetrad (Figure 1(d)).

Single base chain reversal

Chain reversal involves a single adenine, A5, within individual hairpin subunits (Figure 8(a)). The A5 base is bracketed by a closing G6·C4 mismatch involving pairing between the amino group of C4 and the acceptor atoms along the major groove edge of G6 (Figure 2(b)). This concept of a single base hairpin loop bracketed by a mismatch pair, was first reported for single base loops closed by sheared G·A mismatches in antiparallel duplexes.[29,30] There is no pairing between A5 residues across from each other (Figure 9(a)) in the head-to-head hairpin dimeric d(GAGCAGGT) quadruplex. Rather, the A5 residues are stacked over the adjacent C4 residues (Figure 9(a)).

We have checked for non-standard backbone torsion and sugar pucker pseudo-rotation (P) angles within the single residue edgewise turn spanning the C4-A5-G6 segment amongst the refined structures of the dimeric d(GAGCAGGT) quadruplex. For the C4-A5 step, the P value for C4

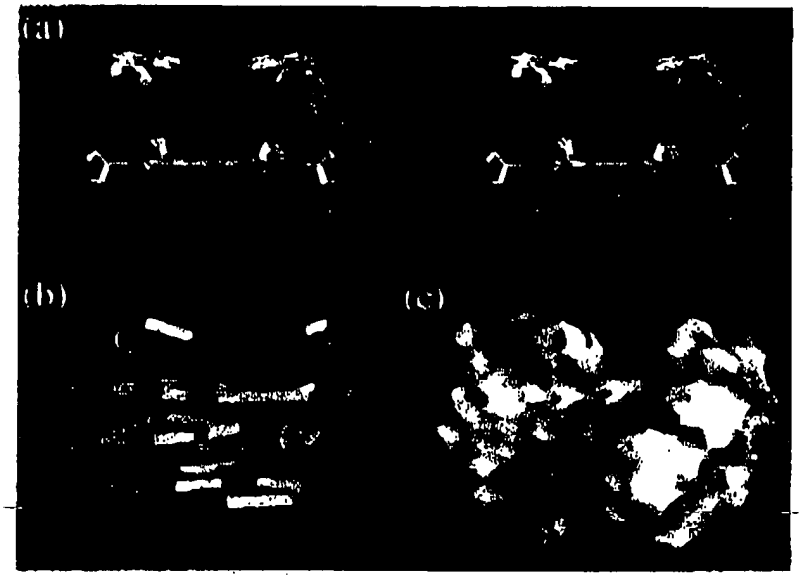

Figure 7. (a) Superpositioned stereo view of ten intensity refined structures of the dimeric d(GAGCAGGT) quadruplex. The backbones of the individual strands are in orange and green with phosphate oxygen atoms removed for clarity. The A5 residue is in white, the G6·C4·G6·C4 tetrad in magenta, G3·G7·G3·G7 tetrad in yellow and A2·T8·A2·T8 tetrad in cyan. (b) A GRASP ribbon view of a representative intensity refined structure of the d(GAGCAGGT) quadruplex. The backbones of individual strands are colored in blue and red. Guanine, adenine and thymine bases are colored yellow, red and blue, respectively. (c) A GRASP surface view of a representative intensity refined structure of the d(GAGCAGGT) quadruplex.

of 69.8(±5.7)° places it in the C4′-*exo* range, the γ value of 149.9(±40.1)° places it in the *trans* range (in contrast to standard *gauche*$^+$ value of 64°), and the χ value for A5 of 170.1(±4.6)° places it in the low *anti* range.

Base stacking

There is excellent pyrimidine/purine and purine/purine stacking between the bases of the direct G6·C4·G6·C4 tetrad (in magenta) and the G3·G7·G3·G7 tetrad (in yellow) in the folded topology of the d(GAGCAGGT) quadruplex (Figure 9(b)). We observe partial stacking between the bases of the G3·G7·G3·G7 tetrad (in yellow) and the slipped A2·T8·A2·T8 tetrad (in cyan) in the folded topology of the d(GAGCAGGT) quadruplex (Figure 9(c)). Overall, there is significant stacking of bases along the length of the dimeric d(GAGCAGGT) quadruplex.

The stacking pattern between the G3·G7·G3·G7 tetrad (in yellow) and the slipped A2·T8·A2·T8 tetrad (in cyan) (Figure 9(c)), positions the imino proton of G7 of one strand in close proximity to the NH_2 and H2 protons of A2 of the partner strand, consistent with the observed inter-strand NOEs (peaks 3 and 3′, Figure 6(a); peak 4, Figure 6(b)) in mixed labeling experiments.

Inter-strand NOEs

We have observed a set of weaker inter-strand NOEs, in addition to their stronger counterparts, shown in Figure 6(a) and (b), that merit further discussion. The studies presented in Figure 6 on an equimolar sample of unlabeled and uniformly ^{13}C,^{15}N-labeled d(GAGCAGGT) quadruplex in 1 M NaCl-containing H_2O buffer, were recorded at 0 °C. We have also collected a ^{15}N-edited (ω_1), ^{13}C,^{15}N-purged (ω_2) NOESY[26] (100 ms mixing time) spectrum, with the carrier on the imino-^{15}N, recorded at 10 °C. This spectrum had improved signal-to-noise, permitting data presentation at a lower contour level (as shown in Supplementary Material, Figure S7). We observe a set of weak inter-strand NOE cross-peaks labeled 5 to 9 in Supplementary Material, Figure S7, in addition to stronger peaks 1 to 4, that were also seen in Figure 6(b).

Peaks 5 and 5′, assigned to inter-strand NOEs between the imino proton of G7 and the amino protons of C4, are consistent with the stacking pattern shown in Figure 9(b). Peaks 6 and 6′, assigned to inter-strand NOEs between the imino proton of G7 and the amino protons of A2, are consistent with the stacking pattern shown in Figure 9(c). Peak 9, assigned to an inter-strand NOE between the imino of G6 and the H2 of A5, is consistent with th stacking pattern shown in Figure 9(a).

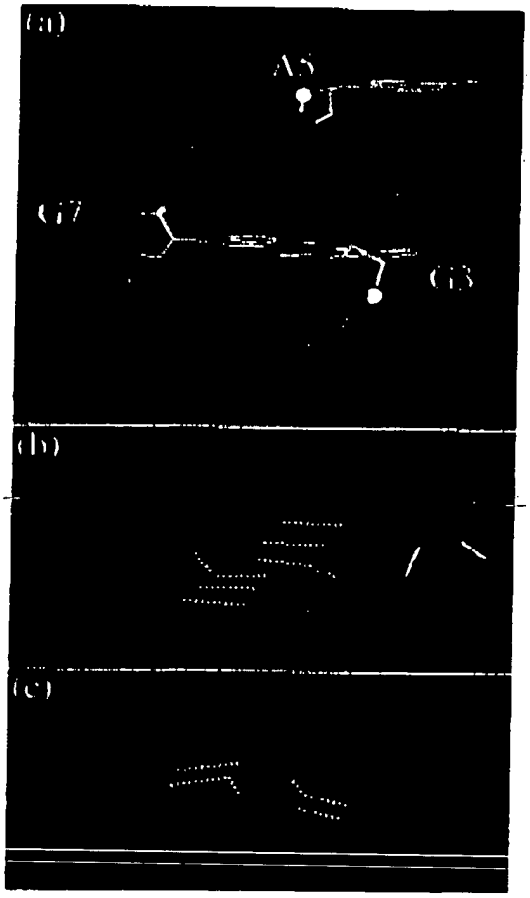

Figure 8. (a) A stick representation of a symmetric half (corresponding to one of the two strands) of a representative intensity refined structure of the d(GAGCAGGT) quadruplex. The color code is the same as outlined in the legend to Figure 7. (b) Pairing alignment of the G6·C4·G6·C4 tetrad. (c) Pairing alignment of the A2·T8·A2·T8 tetrad.

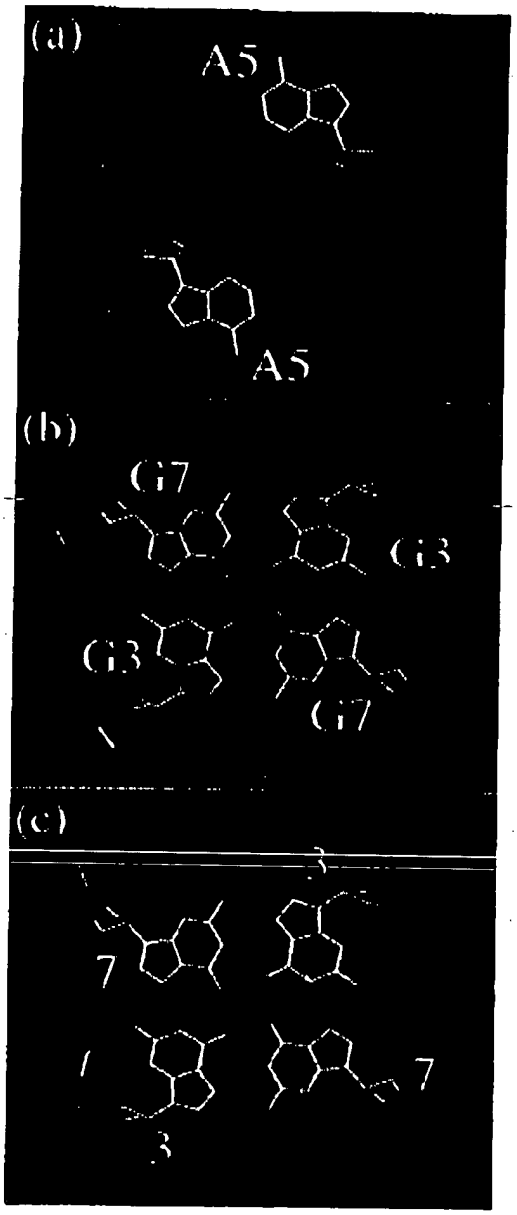

Figure 9. Base stacking overlap patterns in a representative intensity refined structure of the d(GAGCAGGT) quadruplex. (a) Stacking of A5 (in white) on the G6·C4·G6·C4 tetrad (in magenta). (b) Stacking of the G6·C4·G6·C4 tetrad (in magenta) on the G3·G7·G3·G7 tetrad (in yellow). (c) Stacking of the G3·G7·G3·G7 tetrad (in yellow) on the A2·T8·A2·T8 tetrad (in cyan).

By contrast, peak 7 assigned to an inter-strand NOE between the imino of G3 and the H8 of A2, cannot be explained by the overlap pattern in Figure 9(c); Peak 8, representing an inter-strand NOE assigned to the imino of G3 and the H8 proton of G6/G7, cannot be explained by the overlap pattern shown in Figure 9(b). In both cases, while the inter-strand distances between proton pairs are long, their intra-strand counterparts are in close proximity. Thus, it could be argued that perhaps peaks 7 and 8 reflect relatively strong intra-strand NOEs that are not completely removed by isotope filtering. For this explanation to be valid, peaks 7 and 8 should have been doublets, because ^{13}C decoupling was not acquired during acquisition. Currently, we are somewhat at a loss to account for this discrepancy involving weak peaks 7 and 8 (Supplementary Material, Figure S7), whose inten-

sity did vary in the weak range when the experiments were checked for reproducibility.

Salt dependence

The broad resonances for the d(GAGCAGGT) sequence in low NaCl solution (Supplementary Material, Figure S3(a)) is indicative of aggregate formation. What was unexpected was the transition to narrow resonances in 1 M NaCl solution (Supplementary Material, Figure S3(b)), resulting in the formation of a dimeric d(GAGCAGGT) quadruplex, amenable to structural characterization. Both the direct G·C·G·C (Figure 1(a)) and slipped A·T·A·T (Figure 1(d)) tetrads have inwardly pointing amino groups, an alignment not conducive to coordinating monovalent cations positioned either between or within tetrad planes. Therefore, it is conceivable that 1 M NaCl has minimal effect on the integrity of the dimeric d(GAGCAGGT) quadruplex, but for some unknown reason may destabilize the aggregated species that predominates in low salt solution.

Comparison with mixed tetrads involving minor groove alignment of G·C and A·T pairs

Our structural studies on major groove-aligned direct G·C·G·C and slipped A·T·A·T tetrads in the head-to-head dimeric d(GAGCAGGT) quadruplex (this study) can be compared with reported structural studies on quadruplexes involving minor groove-aligned direct G·C·G·C (Supplementary Material, Figure S1(a)) and slipped cation-coordianted A·T·A·T (Supplementary Material, Figure S1(b)) tetrads formed by dimerization of linear and cyclic octameric sequences.[8,9,14] Interestingly, the minor groove-aligned direct G·C·G·C (Supplementary Material, Figure S1(a)) and slipped cation-coordinated A·T·A·T (Supplementary Material, Figure S1(b)) tetrads also involve Watson-Crick G·C and A·T pairing between rather than within head-to-head hairpin dimers of the quadruplex.

It should be noted that while the major groove-aligned direct G·C·G·C (Figure 1(a)) and slipped A·T·A·T (Figure 1(d)) tetrads adopt planar geometries in the solution structure of the dimeric d(GAGCAGGT) quadruplex (this study), the minor groove-aligned direct G·C·G·C (Supplementary Material, Figure S2(a)) and slipped cation-coordianted A·T·A·T (Supplementary Material, Figure S2(b)) tetrads adopt distinctly non-planar geometries within their quadruplexes, both in the crystalline[8,14] and solution[9] states.

Biological relevance

We have reported the first experimental demonstration for formation of a major groove-aligned slipped A·T·A·T tetrad (Figure 1(d)) within a DNA quadruplex architecture. This result, together with our previous demonstration of major groove-aligned direct G·C·G·C tetrad (Figure 1(a))[7] formation within a quadruplex architecture, expands on the code for alignment of duplex segments that can potentially participate in strand exchange during homologous recombination. The glycosidic bonds are directed outwards in major groove aligned tetrads (Figure 1), while they are directed inwards in minor groove-aligned tetrads (Supplementary Material, Figure S1). There is steric crowding between the inwardly pointing sugars in the minor groove-aligned tetrads, resulting in base-sugar and sugar-sugar interactions which severely buckle the planes of the tetrads.[8,9,13] By contrast, there is no such crowding between the outwardly pointing sugars in the major groove-aligned mixed tetrads, resulting in a planar arrangement of bases within the tetrad. Several challenges remain despite our identification here of a major groove-aligned slipped A·T·A·T tetrad (Figure 1(d)) at one end of a quadruplex. Can one identify formation of a major groove-aligned direct A·T·A·T tetrad (Figure 1(c)), and can either the direct or slipped A·T·A·T tetrads be positioned in the interior of the quadruplex?

McGavin[31] was the first to recognize the potential for self-pairing of Watson-Crick G·C and A·T pairs to form dyad axes related major groove-aligned G·C·G·C and A·T·A·T tetrads. Such tetrad-based quadruplexes could play a role in strand exchange between two homologous duplexes during genetic recombination. Wilson[32] next proposed a quadruplex-based model for formation of reciprocal heteroduplexes from their homologous duplex counterparts. In this model, homologous duplexes initially associate through G·C·G·C and A·T·A·T tetrad formation along their minor groove edges, placing complementary strands opposite each other. Strand exchange can then occur following a 90° rotation of each base, such that the base-pairs now face each other through G·C·G·C and A·T·A·T tetrad formation along their major groove edges. Experimental studies[33,34] have provided evidence for quadruplex formation by poly(CA)$_n$·poly(TG)$_n$ repeats, while computational studies[35] have established that such strand exchange quadruplexes, as well as quadruplex-duplex junctions, are stereochemically robust and form stable entities. Thus, a future challenge would be to design a DNA quadruplex consisting solely of G·C·G·C and A·T·A·T tetrads and attempt to trap structures where the alignment is through the minor groove on the one hand, and through the major groove on the other.

Materials and Methods

Preparation of unlabeled and uniformly ^{13}C, ^{15}N-labeled DNA

The unlabeled d(GAGCAGGT) sequence was synthesized on a 10 μmol scale on an Applied Biosystems 392 DNA synthesizer using solid phase β-cyanoethyl-

phosphoamidite chemistry and was subsequently purified by high pressure liquid chromatography (HPLC).

A modified version of the Zimmer & Crothers[3a] procedure as described[19,26] was used for the enzymatic synthesis of uniformly ^{13}C,^{15}N-labeled d(GAGCAGGT) sequence. The in-house prepared uniformly ^{13}C,^{15}N-labeled dNTPs[19,26] were used as building blocks in the *in vitro* polymerization reaction catalyzed by murine mammary leukemia virus (MMLV) reverse transcriptase (Gibco-BRL). The uniformly ^{13}C,^{15}N-labeled d(GAGCAGGT) 8-mer was separated from the unlabeled 24-mer template using 22% (w/v) denaturing polyacrylamide electrophoresis. The DNA 8-mer bands were eluted from the gel by "crush and soak" procedure and purified as described above for the non-labeled samples.

NMR data collection and processing

NMR data on the d(GAGCAGGT) 8-mer in H_2O and ^2H_2O buffer (1 M NaCl, 2 mM phosphate (pH 6.6)) were collected on a Varian 600 MHz Unity *Inova* NMR spectrometer. Proton assignments are based on homonuclear NOESY, correlation spectroscopy (COSY), TOCSY and HCCNH-TOCSY experiments. Data sets were processed and analyzed using the FELIX program (Molecular Simulations).

Two bond $^{2h}J_{NN}$ scalar couplings between imino and amino donors and nitrogen acceptors in uniformly ^{13}C,^{15}N-labeled d(GAGCAGGT) in 1 M NaCl were monitored in HNN-COSY[21,22] and H(CN)N(H)[24,25] contour plots using pulse sequences described in the literature.

Distance restraints

The distances between non-exchangeable protons were estimated from the buildup curves of cross-peak intensities in NOESY spectra at four different mixing times (50, 150, 200 and 250 ms) in ^2H_2O and given bounds of ±30%, with distances referenced relative to the cytosine H6-H5 distance of 2.47 Å. Exchangeable proton restraints are based on NOESY data sets at two mixing times (60 and 200 ms) in H_2O. Cross-peaks involving exchangeable protons were classified as strong (medium to strong intensity at 60 ms), medium (weak intensity at 60 ms) and weak (observed only at a mixing time of 200 ms) and proton pairs were then restrained, respectively, to distances of 3.0(±0.9) Å, 4.0(±1.2) Å and 6.0(±1.8) Å. Since the experimental NMR data are consistent with a dimeric quadruplex motif, non-crystallographic symmetry restraints were imposed on all heavy atoms.

Structure calculations

The structure of the d(GAGCAGGT) in 1 M NaCl was determined by molecular dynamics (MD)-simulated annealing computations driven by NOE distance and hydrogen bonding restraints using the X-PLOR package, version 3.8.[27] At the initial stage of the refinement, torsional molecular dynamics was undertaken at high temperature. The molecules were equilibrated at 20,000 K (30,000 steps over 3 ps) and then cooled very slowly to 1000 K (40,000 steps over 20 ps). The potential energy function included a repulsive force field, NOE and hydrogen bond distance restraints, glycosidic bond (χ) dihedral angle restraints and a non-crystallographic symmetry potential. The force constant for NOE distance restraints was maintained at a value of 30 kcal mol^{-1} Å^{-2}, while for hydrogen bonds restraints the value was 50 kcal mol^{-1} Å^{-2}. All NOE distance restraints were considered as ambiguous and treated with the "sum" averaging option.[28,29] Dihedral angle restraints (210(±40)°, with force constant of 50 kcal mol^{-1} rad^{-2}) were imposed on glycosidic torsion angles for the residues A2, C4, A5, G6, G7 and T8 shown experimentally to adopt *anti* conformations. Dihedral angle restraints (65(±40)°, with force constant of 50 kcal mol^{-1} rad^{-2}) were imposed on glycosidic torsion angles for the G3 residue shown experimentally to adopt *syn* conformation. The force constant for non-crystallographic symmetry was maintained at 30 kcal mol^{-1} Å^{-2}.

These computations were followed by lower temperature Cartesian space molecular dynamics guided by the hydrogen bonding and NOE distance restraints with changes in the potential energy function: the repulsive force field was replaced with Lennard-Jones potentials and planarity restraints were included for tetrad planes with low weights of 5 kcal mol^{-1} Å^{-2} and 10 kcal mol^{-1} Å^{-2}, respectively. During this stage of the dynamics, the structures were further cooled from 1000 K to 300 K (20,000 steps over 10 ps) and minimized until the gradient of energy was less than 0.1 kcal mol^{-1}. It should be noted that computations repeated without planarity restraints resulted in the same low energy structures, but exhibited a lower convergence rate.

The refinement protocol started from 60 different initial structures. The initial structures were generated as sets of two chains, each eight nucleotides long, randomized for all dihedral angles, and separated by space intervals of 50 Å. The convergence rate following dynamics was good for the case where the computations were guided by hydrogen bonding restraints associated with the topology shown in Figure 2(a): Sixteen structures out of 60 emerged with the same fold and pairwise r.m.s.d. values less than 1.0 Å between members of the group. Non-converged structures were separated from that group by large gaps (in total more than 400 kcal) in all components of the potential energy (van der Waals, NOE violations, covalent geometry).

The ten converged distance refined structures corresponding to the folding topology shown schematically in Figure 1(a) were used as the starting point for subsequent X-PLOR based energy minimization with back-calculation of the NOESY spectra. The relaxation matrix was set up for the non-exchangeable protons, with the exchangeable imino and amino protons exchanged for deuterons. A total of 712 non-exchangeable intensity values from NOESY data sets at four mixing times in ^2H_2O buffer (178 non-exchangeable intensities per mixing time) were included with force constant of 500 kcal mol^{-1}. The planarity restraints were set to very low values of 1.0 kcal Å^{-2} for the G·G·G·G tetrad and 0.5 kcal Å^{-2} for the A·T·A·T and G·C·G·C tetrads. The distance restraints were retained with 30% bounds and the same weights as before. During minimization, the NMR R factor ($R_{1/\rho}$) improved from the initial value of 12% to 7.4% while retaining structure convergence and stereochemistry.

Coordinates deposition

Coordinates (accession number: 1jvc) of the dimeric d(GAGCAGGT) quadruplex have been deposited in the RCSB Protein Data Bank.

Acknowledgments

This research was supported by NIH grants GM-34504 to D.J.P.

References

1. Williamson, J. R. (1994). G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703-730.
2. Rhodes, D. & Giraldo, R. (1995). Telomere structure and function. *Curr. Opin. Struct. Biol.* **5**, 311-312.
3. Feigon, J., Koshlap, K. M. & Smith, F. W. (1995). ^{1}H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.* **261**, 225-255.
4. Patel, D. J., Bouaziz, S., Kettani, A. & Wang, Y. (1999). Structures of guanine-rich and cytosine-rich quadruplexes formed *in vitro* by telomeric, centromeric and triplet repeat disease sequences. *Oxford Handbook of Nucleic Acid Structures* (Neidle, S., ed.), pp. 389-453, Oxford University Press, Oxford.
5. Arthanari, H. & Bolton, P. H. (2001). Functional and dysfunctional roles of quadruplex DNA in cells. *Chem. Biol.* **8**, 221-230.
6. Sen, D. & Gilbert, W. (1991). The structure of telomeric DNA: DNA quadruplex formation. *Curr. Opin. Struct. Biol.* **1**, 435-438.
7. Kettani, A., Kumar, R. A. & Patel, D. J. (1995). Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.* **254**, 636-656.
8. Leonard, G. A., Zhang, S., Peterson, M. R., Harrop, S. J., Helliwell, J. R., Cruse, W. B. *et al.* (1995). Self-association of a DNA loop creates a quadruplex crystal structure of d(GCATGCT) at 1.8 Å resolution. *Structure*, **3**, 335-340.
9. Escaja, N., Fedroso, E., Rico, M. & Gonzalez, C. (2000). Dimeric solution structure of two cyclic octamers. Four-stranded DNA structures stabilized by A·T·A·T and G·C·G·C tetrads. *J. Am. Chem. Soc.* **122**, 12732-12742.
10. Kettani, A., Bouaziz, S., Gorin, A., Zhao, H., Jones, R. & Patel, D. J. (1998). Solution structure of a Na$^+$ cation stabilized DNA quadruplex containing G·G·G·G and G·C·G·C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.* **282**, 619-636.
11. Bouaziz, S., Kettani, A., Zhao, H., Jones, R. & Patel, D. J. (1998). A K$^+$ cation dependent conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats. *J. Mol. Biol.* **282**, 637-652.
12. Majumdar, A., Kettani, A., Skripkin, E. & Patel, D. J. (2001). Pulse sequences for detection of $NH_2\cdots N$ hydrogen bonds in a sheared G·A mismatches *via* remote, nonexchangeable protons. *J. Biomol. NMR*, **19**, 103-113.
13. Salisbury, S. A., Wilson, S. E., Powell, H. R., Kennard, O., Lubini, P., Sheldrick, G. M. *et al.* (1997). The bi-loop, a new general four-stranded DNA motif. *Proc. Natl Acad. Sci. USA*, **94**, 5515-5518.
14. Kuryavyi, V. V. & Jovin, T. M. (1995). Triad-DNA: a model for trinucleotide repeats. *Nature Genet.* **9**, 339-341.
15. Kettani, A., Bouaziz, S., Wang, W., Jones, R. A. & Patel, D. J. (1997). *Bombyx mori* single repeat telomeric DNA sequences forms a G-quadruplex capped by base triads. *Nature Struct. Biol.* **4**, 382-389.
16. Kettani, A., Basu, G., Gorin, A., Majumdar, A., Skripkin, E. & Patel, D. J. (2000). A two-stranded template-based approach to G·(C-A) triad formation: designing novel structural elements into an existing DNA framework. *J. Mol. Biol.* **301**, 129-146.
17. Kuryavyi, V., Kettani, A., Wang, W., Jones, R. & Patel, D. J. (2000). A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads. *J. Mol. Biol.* **295**, 455-469.
18. Kuryavyi, V., Majumdar, A., Shallop, A., Chernichenko, N., Skripkin, E., Jones, R. A. & Patel, D. J. (2001). A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G(*syn*)·G(*syn*)·G(*anti*)·G(*anti*) tetrads flanked by a G·(T-T) triad and a T·T·T triple. *J. Mol. Biol.* **310**, 181-194.
19. Kettani, A., Gorin, A., Majumdar, A., Hermann, T., Skripkin, E., Zhao, H. *et al.* (2000). A dimeric DNA interface stabilized by A·(G·G·G·G)·A hexads and coordinated monovalent cations. *J. Mol. Biol.* **297**, 627-644.
20. Phan, A. T. (2000). Long range imino proton-^{13}C J couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR*, **16**, 175-178.
21. Dingley, A. J. & Grzesiek, S. (1998). Direct observation of hydrogen bonds in nucleic acid base-pairs by internucleotide $^2J_{NN}$ couplings. *J. Am. Chem. Soc.* **120**, 8293-8297.
22. Pervushin, K., Ono, A., Fernandez, C., Szyperski, T., Kainosho, M. & Wuthrich, K. (1998). NMR scalar couplings across Watson-Crick base-pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proc. Natl Acad. Sci. USA*, **95**, 14147-14151.
23. Majumdar, A., Kettani, A. & Skripkin, E. (1999). Observation and measurement of $^2J_{NN}$ coupling constants between ^{15}N nuclei with widely separaterd chemical shifts. *J. Biomol. NMR*, **14**, 67-70.
24. Majumdar, A., Kettani, A. & Skripkin, E. (1999). Observation of internucleotide N-H··N hydrogen bonds in the absence of directly detectable protons. *J. Biomol. NMR*, **15**, 207-211.
25. Hennig, M. & Williamson, J. R. (2000). Detection of N-H··N hydrogen bonding in RNA via scalar couplings in the absence of observable imino proton resonances. *Nucl. Acids Res.* **28**, 1585-1593.
26. Kettani, A., Bouaziz, S., Skripkin, E., Majumdar, A., Wang, W., Jones, R. A. & Patel, D. J. (1999). Interlocked mismatch-aligned arrowhead DNA motifs. *Structure*, **7**, 803-815.
27. Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A. & Feigon, J. (1993). Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl Acad. Sci. USA*, **90**, 3745-3749.
28. Wang, K. Y., Krawczyk, S. H., Bischofberger, N., Swaminathan, S. & Bolton, P. H. (1993). The tertiary structure of a DNA aptamer which binds to and inhibits thrombin determines activity. *Biochemistry*, **32**, 11285-11292.
29. Chou, S. H., Zhu, L., Gao, Z., Cheng, J-W. & Reid, B. R. (1996). Hairpins loops consisting of single adenine residues closed by a sheared A·A and G·G pairs formed by the DNA triplets AAA and GAG:

Solution structure of the d(GTACAAAGTAC). *J. Mol. Biol.* **264**, 981-1001.
30. Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K. & Miura, K. (1994). Most compact hairpin turn structure exerted by a short DNA fragment d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat. *Nucl. Acids Res.* **22**, 576-582.
31. McGavin, S. (1971). Models of specifically paired like (homologous) nucleic acid structures. *J. Mol. Biol.* **55**, 293-298.
32. Wilson, J. H. (1979). Nicked-free formation of reciprocal heteroduplexes: a simple solution to the topological problem. *Proc. Natl Acad. Sci. USA*, **76**, 3641-3645.
33. Johnson, D. & Morgan, A. R. (1978). Unique structures formed by pyrimidine·purine DNA which may be four-stranded. *Proc. Natl Acad. Sci. USA*, **75**, 1637-1641.
34. Gaillard, C. & Strauss, F. (1994). Association of poly(CA)·poly(TG) DNA fragments into four-stranded complexes bound by HMG1 and 2. *Science*, **264**, 433-436.
35. Lebrun, A. & Lavery, R. (1995). Modeling a strand exchange tetraplex conformation. *J. Biomol. Struct. Dynam.* **13**, 459-464.
36. Zimmer, D. P. & Crothers, D. M. (1995). NMR of enzymatically synthesized uniformly ^{13}C, ^{15}N-labeled DNA oligonucleotides. *Proc. Natl Acad. Sci. USA*, **92**, 3091-3095.
37. Brunger, A. T. (1992). *X-PLOR. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
38. Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE cross-peaks and disulfide connectivities. *J. Mol. Biol.* **245**, 645-660.
39. Nilges, M., Habazettl, J., Brunger, A. T. & Holak, T. A. (1991). Relaxation matrix alignment of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.* **219**, 499-510.

Edited by M. F. Summers

(Received 22 May 2001; received in revised form 8 August 2001; accepted 8 August 2001)

http://www.academicpress.com/jmb

Supplementary Material for this paper comprising seven Figures and two Tables are available on IDEAL